At Brookhaven, lattice quantum chromodynamics calculations have been critical in understanding the results of particle and nuclear physics experiments to determine how quarks and gluons interact, and models of materials have provided insight into their behaviors and properties, especially at the nanoscale.

The role of computational science in the scientific discovery process will significantly expand as exascale computing—the ability to compute at a rate of at least 1018 (a billion billion) floating-point operations per second (exaFLOPS)—becomes a reality, enabling scientists to solve larger and more complex problems. But realizing the full benefits of computing at the exascale will in part require advances in numerical algorithms.

At Brookhaven's Computational Science Laboratory (CSL), led by computer scientist Nicholas D'Imperio, experts in HPC, applied mathematics, and domain science are adapting, optimizing, and developing such algorithms. These exascale-ready algorithms will help solve problems in computational physics, biology, chemistry, materials science, and environmental science. Current projects at Brookhaven include the development of an exascale-compatible version of the computational chemistry code NWChem for applications related to biomass-based energy production, and of methods to accurately describe the electronic properties of high-temperature super-conductors and other strongly correlated materials.

As exascale computing becomes a reality, CSL is also looking at other novel computer architectures and exploring how they can be effectively used to model and predict the physical world in ever-greater detail and with increased accuracy, within the confines of the available computing capabilities. Areas of particular emphasis in CSL's work include the support of data-intensive applications, workflows with specific performance requirements, and parallelization techniques involving graphics processing unit (GPU), Intel Phi, and field-programmable gate array accelerators.

Center for Data-Driven Discovery

A centerpiece of Brookhaven's CSI is the Center for Data-Driven Discovery (C3D), currently led by CSI Director Kerstin Kleese van Dam. C3D is driving the research and development of new methods, tools, and services to extract knowledge and discover new scientific insights from big data.

The center builds on the Laboratory's leadership in scientific data management and analysis and facilitates collaborations between CSI and Brookhaven's directorates and external partners to fulfill four core missions:

- Provide a focal point for leading data science research

- Develop, test, and provide cyber infrastructure and real-time data-processing and decision-making tools for next-generation big data projects, including new NSLS-II beamlines and the proposed new Electron-Ion Collider

- Educate the next generation of expert data scientists

- Translate research advances into tools and expertise that lead to measurable scientific progress and improved industrial competitiveness, driving collaborations with Brookhaven's research partners and experimental facility users, such as General Electric and Pfizer

Hundreds of scientists from Brookhaven and thousands of facility users from universities, industry, and other laboratories around the country and throughout the world will benefit from the capabilities that C3D personnel develop to analyze, simulate, and predict complex phenomena important to many areas of science.

The goals of C3D and CSI overall are well aligned with the broad missions of many organizations, especially those of DOE's Office of Science.

Strategic Partnerships

CSI seeks partnerships with key players in academia, government, and industry to advance scientific research and offer its expertise and tools to the community.

One particularly successful example of a partnership is the collaboration between Brookhaven, Columbia University, IBM, and the University of Edinburgh that in the early 2000s led to the development of the IBM Blue Gene supercomputing architecture, which continues to be used on the world's most powerful commercially available supercomputers.

More recent partnerships include those with other national laboratories and leading universities to work on DOE's Exascale Computing Project; with local New York State–based companies, such as Reservoir Labs and Kitware, through DOE's Small Business Innovation Research program; and with PayPal and NVIDIA, which named Brookhaven an NVIDIA GPU Research Center in 2016.

Going forward, CSI is particularly interested in expanding its partnerships with industry, where its research, tools, and infrastructure services could have a profound impact. Existing efforts in modeling and data analytics for sectors such as electric power and health and life sciences are excellent foundations for future collaborations.

We look forward to hearing from you and working together to drive discovery and innovation.

Contacts

Kerstin Kleese van Dam
Director, Computational Science Initiative
phone: 631-344-6019, email: kleese@bnl.gov

Lauri Peragine
Administrative Specialist, Computational Science Initiative
Phone: 631-344-7090, email: lperagine@bnl.gov

https://www.bnl.gov/compsci/

U.S. DEPARTMENT OF ENERGY
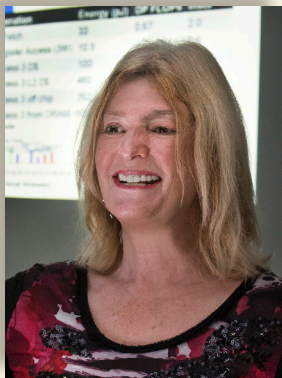
BROOKHAVEN
NATIONAL LABORATORY

Brookhaven National Laboratory
Computational Science Initiative
Leveraging Expertise and Investments to Translate Big Data into Discovery and Innovation

Kerstin Kleese van Dam, director of Brookhaven Lab's Computational Sciences Initiative and interim director of the Center for Data-Driven Discovery (C3D)

Frank Alexander, deputy director of Brookhaven Lab's Computational Science Initiative

Barbara Chapman, chair of the Computer Science and Mathematics Department at Brookhaven Lab and professor of applied mathematics and statistics and computer science at Stony Brook University

Computational Science Initiative
Leveraging Expertise and Investments to Translate Big Data into Discovery and Innovation

The Computational Science Initiative (CSI) integrates computer science, applied mathematics, computational science, and domain science expertise and investments across the U.S. Department of Energy's (DOE) Brookhaven National Laboratory—including its flagship facilities that attract thousands of scientific users each year—to tackle the big data challenges at the frontiers of scientific discovery.

The timely analysis and interpretation of data are essential to addressing the nation's grand challenges in science, national security, and industry. In particular, CSI is investigating novel approaches to harnessing the power of large data volumes and data rates to drive discovery. To this end, CSI pursues the research and development of enabling and translational capabilities.

Enabling Capabilities
- The Computer Science and Mathematics Department, conducting research into cutting-edge methodologies in support of large-scale, multimodal, and streaming data analysis
- The Scientific Data and Computing Center, housing the latest systems in high-throughput, high-performance, and data-intensive computing, data storage, and networking, and offering everything from innovative research platforms to highly reliable production services

Translational Capabilities
- The Computational Science Laboratory, a collaborative center for the development, characterization, and optimization of advanced algorithms
- The Center for Data-Driven Discovery, a multidisciplinary center for the development, deployment, and operation of data-intensive discovery services specifically tailored for science, national security, and industry

The CSI Philosophy
The CSI takes a multidisciplinary and collaborative approach to scientific research and development, with activities targeted at and informed by the key challenges observed in close interactions with our clients in science, national security, and industry. Our success is measured in equal parts by the advancements we can bring to computer science and applied mathematics and by the transformational impacts we have on our clients' mission areas.

The CSI brings together under one umbrella the subject-matter expertise needed to efficiently and effectively transform scientific big data into meaningful insights, making optimal use of existing technologies while also leading the development of new tools and methods that will accelerate discovery both within and beyond the Laboratory. Key partners include IBM, Intel, Kitware, NVIDIA, and nearby universities such as Columbia, Cornell, NYU, Rutgers, and Stony Brook.

Computer Science and Mathematics Department
The Computer Science and Mathematics Department—led by Barbara Chapman, a professor of applied mathematics and statistics and computer science at Stony Brook University—primarily focuses on the development of new methods and approaches to make sense of data at the exabyte scale and beyond.

The real-time analysis of ultra-high-throughput data streams, such as those produced by in situ and in operando experiments at large-scale scientific research facilities, is important to optimizing experimental design and producing high-quality results. To enable such analysis, the department develops integrated extreme-scale machine learning and visual analytics methods, on-the-wire processing algorithms and networking infrastructure, and data-intensive scientific workflows.

In addition, the department applies theory and computational models and simulations to the autonomous design and execution of experiments. Automating this part of the scientific process could improve the accuracy and reproducibility of experiments and increase the rate of discovery.

Also critical to accelerating discovery is the ability to interactively explore multi-petabyte datasets, particularly those of nuclear and high-energy physics, computational biology, and climate science. The department performs research into the novel hardware, software, programming models, and visual analytics paradigms needed to support such interactive explorations.

Scientific Data and Computing Center
Brookhaven Lab supports data-intensive computing and analysis requirements at the extreme scale. In 2017, Brookhaven exceeded 100 petabytes (PB) of archived data, becoming the second largest scientific data archive in the nation and the fourth largest worldwide. In supporting key DOE Office of Science User Facilities—the Relativistic Heavy Ion Collider (RHIC), the National Synchrotron Light Source II (NSLS-II), and the Center for Functional Nanomaterials (CFN)—as well as the international ATLAS experiment at CERN's (the European Organization for Nuclear Research) Large Hadron Collider (LHC), Brookhaven analyzed more than 400 PB of scientific results in 2016, and expects to exceed 500 PB in 2017.

Data storage and processing at such scales are enabled by an integrated data management, computing, and networking infrastructure provided by the Scientific Data and Computing Center (SDCC). An ever-expanding farm of computing nodes receives data from the millions of particle collisions that take place each second at RHIC and the billions at LHC—storing and processing these data and analyzing and distributing them for collaborators throughout the world.

Through its institutional cluster and novel architecture test bed, SDCC provides access to the latest high-performance-computing (HPC) architectures, customized for data-intensive applications. Brookhaven's HPC expertise dates back to its operation of a string of supercomputers named in the TOP500 list, including the New York Blue IBM Blue Gene, which ranked number five in 2007.

This computing provision is rounded out by a Brookhaven cloud service maintained by SDCC. Brookhaven is currently connected to DOE's Energy Sciences Network, which provides advanced networking and communications support to research programs.

This versatile approach to data-intensive computing, combined with new methods for handling data-rich simulations, has helped establish Brookhaven as a leader in high-capacity computing. PanDA (for Production and Distributed Analysis)—a workload management system co-developed by Brookhaven, other DOE national laboratories, and collaborating universities—enabled the analysis of 1.6 exabytes of ATLAS experimental data in 2016. SDCC Director Eric Lançon is leading the development of new key services in this space as part of SDCC's ongoing expansion.

Computational Science Laboratory
Computational modeling and simulation of real-world processes and phenomena play an important role in virtually every field of science.

Nicholas D'Imperio, chair of Brookhaven Lab's Computational Science Laboratory

Eric Lançon, director of Brookhaven Lab's Scientific Data and Computing Center, which includes the RHIC and ATLAS Computing Facility

Robert Harrison, chief scientist of Brookhaven Lab's Computational Science Initiative and director of Stony Brook University's Institute for Advanced Computational Science